平成30年度 情報工学コース卒業研究報告要旨

村瀬勉 研究室	氏　　名	白倉大河
卒業研究題目	クローラ等のHTTPリクエスト自動収集/簡易解析システム reqhack	

コンピュータを相互に接続するインターネットはパソコンやスマートフォンに広く普及し、無くてはならないインフラとなっている。インターネットで重要な地位を占めているプロトコルとして、コンテンツの転送に使われるHTTPと、それを暗号化してやり取りするHTTPSがある。Webブラウザを使用してコンテンツやアプリケーションを利用するWWW(World Wide Web)、サーバクライアント間通信が必要なさまざまなアプリケーションが使用するREST API、双方向通信を実現するWebSocketなどで使われる。

Webサーバにはコンテンツを閲覧するWebブラウザや検索エンジンのクローラなどの正当なアクセスがある一方で、悪意のある攻撃やスキャンにも晒されている。正当なクローラからのアクセスを識別したり、悪意のあるアクセスを調査するためにはログを収集し、解析する必要がある。そのためには調査に必要な情報を収集できるWebサーバ、ドメイン、SSL証明書などを用意しなければならない。また、収集したログを解析する際にもツールを用意したり、HTTPの仕様を確認するといった手間が必要となる。

本研究ではHTTPとHTTPSのリクエストやクライアントの接続情報などを記録し、ブラウザ上から簡便に閲覧可能なシステムreqhackを構築した。システムの構成とデータのやり取りをまとめた図を図1に示す。reqhackの利用方法を図1の番号を使って説明する。まず、解析者はreqhackで解析用URLを作成し①、作成した解析用URLを攻撃者がクロールすると考えられる場所に公開する②。攻撃者が解析用URLにアクセスすると、データベースにその時のリクエストが保存される③。任意の時間の経過後に、解析者は解析結果表示ページを閲覧し④、閲覧時までに収集された攻撃者のリクエストを閲覧し確認する⑤。これにより、既存のWebサーバアプリケーションでログ収集する場合と比べて小さな手間で総合的な記録、解析機能を提供できる。記録の閲覧時にはデータを整形表示したり、簡易的なフィルタリング機能など、解析を補助する機能を提供する。

実際にこのシステムを運用し、悪意のあるボットが閲覧している可能性のあるWebサイトに記録用URLを記載したところ、24時間で13回のリクエストが記録された。解析ページで閲覧できる情報を元に解析を行ったところ、Googleのクローラによる正当なアクセスと機密情報を収集している悪意あるリクエストを確認することができた。

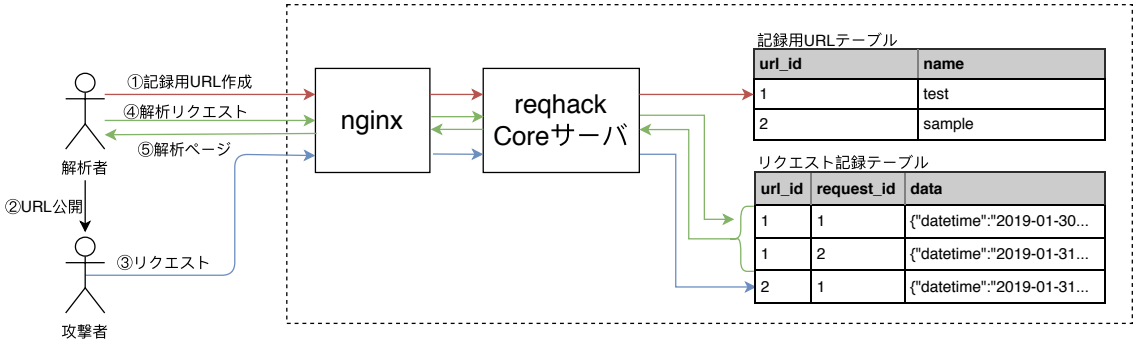

図1. reqhackの構成とデータのやり取り